

**IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MASSACHUSETTS**

SPIDER SEARCH ANALYTICS LLC

Plaintiff,

v.

HUBSPOT, INC.

Defendant.

CIVIL ACTION NO. 17-11285 ()

TRIAL BY JURY DEMANDED

COMPLAINT FOR INFRINGEMENT OF PATENT

COMES NOW, Plaintiff Spider Search Analytics LLC (“SSA” or Plaintiff), through the undersigned attorneys, and respectfully alleges, states, and prays as follows:

NATURE OF THE ACTION

1. This is an action for patent infringement under the Patent Laws of the United States, Title 35 United States Code (“U.S.C.”) to prevent and enjoin defendant Hubspot, Inc., (hereinafter “Defendant”) from infringing and profiting, in an illegal and unauthorized manner and without authorization and/or of the consent from SSA, from U.S. Patent No. 7,454,430 (the “430 patent”, attached hereto as Exhibit “A”) (the “Patent-in-Suit”) pursuant to 35 U.S.C. § 271, and to recover damages, attorney’s fees, and costs.

THE PARTIES

2. Plaintiff SSA is a Texas entity with its principal place of business at 101 E. Park Blvd, Suite 600, Plano, Texas 75074.

3. Upon information and belief, Defendant is a company organized and existing under the laws of the State of Delaware, having a principal place of business at 25 First Street, 2nd Floor,

Cambridge, MA 02141. Upon information and belief, Defendant may be served with process at Brian Halligan 25 First Street, 2nd Floor, Cambridge, MA 02141.

JURISDICTION AND VENUE

4. The Court has subject matter jurisdiction over this action pursuant to 28 U.S.C. §§ 1331 and 1338(a) because the action arises under the Patent Laws of the United States, 35 U.S.C. §§ 1 *et seq.*

5. This Court has personal jurisdiction over Defendant by virtue of its systematic and continuous contacts with this jurisdiction, including having the right to transact business in Massachusetts, as well as because of the injury to SSA, and the cause of action SSA has risen, as alleged herein.

6. Defendant is subject to this Court's personal jurisdiction pursuant to due process, due at least to its substantial business and purposeful availment of this forum, including: (i) at least a portion of the infringements alleged herein; and (ii) regularly doing or soliciting business, engaging in other persistent courses of conduct, and/or deriving substantial revenue from goods and services provided to individuals in Massachusetts and in this judicial district.

7. Upon information and belief, Defendant, directly and/or through its employees or agents, and/or its customers, uses products, as defined below, that perform each and every step of at least one claim of the '430 patent with the knowledge and/or understanding that such products are used or will be used in this District. For example, Defendant uses the accused product to extract data from web pages.¹ Upon information and belief, Defendant has engaged in substantial and not isolated activity within this District. Therefore, exercise of jurisdiction over Defendant will not offend traditional notions of fair play and substantial justice.

¹ <http://www.diffbot.com>.

8. Defendant has conducted and conducts business within the state of Massachusetts, directly or through intermediaries, or offers and advertises products or services, or uses services or products in Massachusetts, including this judicial district, in a manner that infringes the Patents-In-Suit.

9. Specifically, Defendant solicits business from and markets its services to consumers within Massachusetts by providing a service that requires the use of a method for building a deep web crawler, as described in the Patents-In-Suit.

10. In addition to Defendant's continuously and systematically conducting business in Massachusetts, the causes of action against Defendant are connected (but not limited) to Defendant's purposeful acts committed in the state of Massachusetts including Defendant's use of a method for building a deep web crawler, as described in the Patents-In-Suit.

11. Defendant is a company that has a regular and established presence in this district and uses a web crawler that is capable of crawling any website (including, upon information and belief, websites from the "deep web" and that use AJAX for dynamic pages) in order to crawl and scrape data for their clients.

12. Defendant uses scout crawling rules to collect dynamic pages.

13. Defendant uses an analyzer and extractor to determine underlying structure of queries.

14. Defendant generates instructions for a harvester, wherein the harvester provides requests to a server and collects available pages from the server.

15. Venue is proper in this judicial district pursuant to 28 U.S.C. § 1400(b) because Defendant has committed acts of infringement and has a regular and established place of business in this district.

FACTUAL ALLEGATIONS

16. On November 18, 2008, the United States Patent and Trademark Office (“USPTO”) duly and legally issued the ‘430 patent, entitled “System and method for facts extraction and domain knowledge repository creation from unstructured and semi-structured documents” after a full and fair examination. (Exhibit A).

17. SSA is presently the owner of the patent, having received all right, title and interest in and to the ‘430 patent from the previous assignee of record. SSA possesses all rights of recovery under the ‘430 patent, including the exclusive right to recover for past infringement.

18. The ‘430 patent contains three (3) independent claims and twenty-four (24) dependent claims.

19. The ‘430 patent claims, *inter alia*, a method for building a deep web crawler.

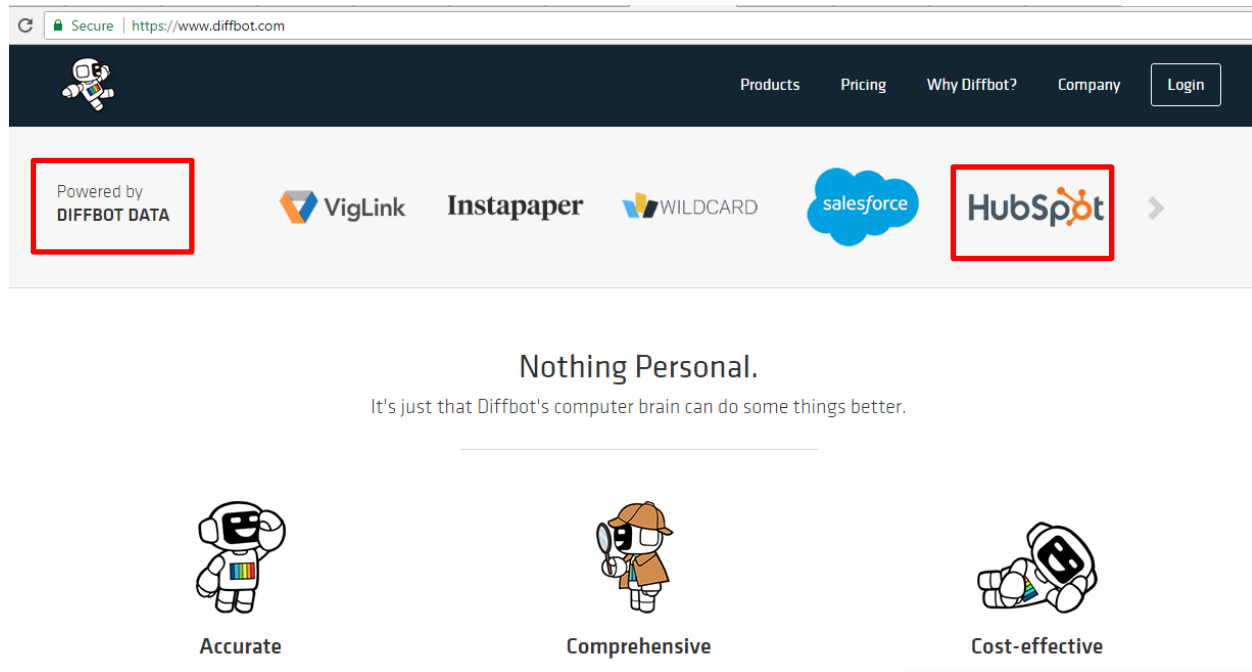
20. Defendant uses, *inter alia*, a method that performs each and every step of at least one claim of the ‘430 patent.

DEFENDANT’S PRODUCTS

21. In accordance with claim 10 of the ‘430 patent, Defendant uses a method for building a deep web crawler. For example, Defendant uses the web crawler “Diffbot” (the “Accused Product”) in order to extract data from web pages.²

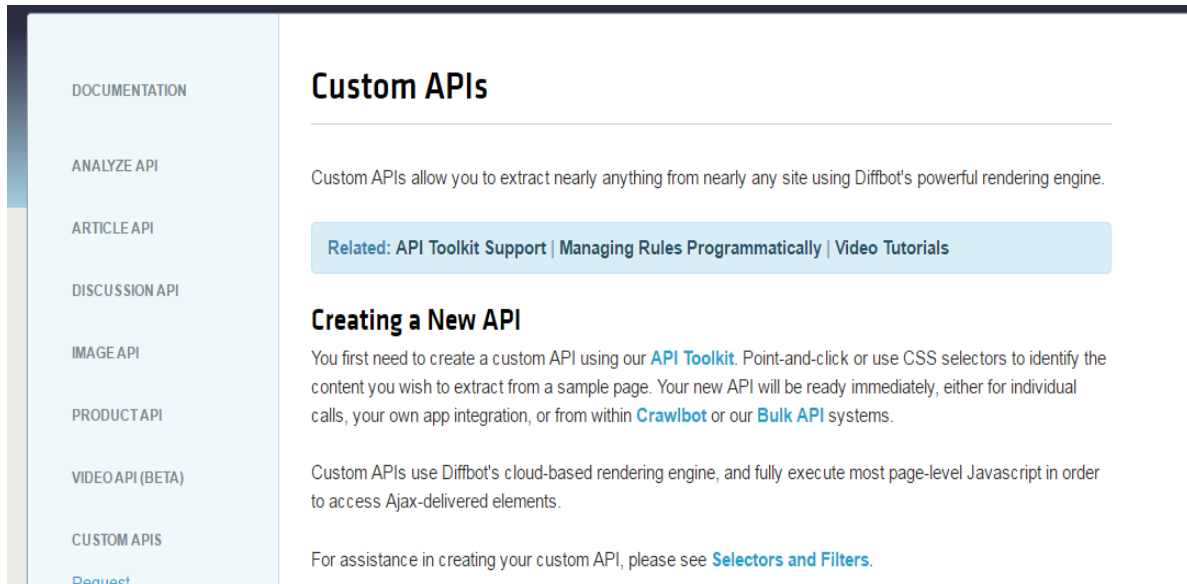
22. As seen below, Defendant is a customer of Diffbot and uses the accused product.

² *Id.*



23. The Accused Product allows customers to build web crawlers that are capable of crawling any website (including, upon information and belief, websites from the “deep web” and that use Javascript and AJAX for dynamic pages).

24. In accordance with claim 10 of the ‘430 patent, Defendant utilizes scout crawling rules to collect dynamic pages.



Source: <https://www.diffbot.com/dev/docs/custom/>

25. In accordance with claim 10 of the '430 patent, Defendant utilizes an analyzer and extractor to determine underlying structure of queries. For example, Accused Product will utilize an Analyze API to determine the "page-type" of submitted URLs, and automatically extract the contents.

Creating a Crawl: Basics

Each crawl requires, at minimum, the following:

1. A crawl name (e.g., "DiffbotCrawl").
2. A starting or "seed" URL. Multiple URLs can be provided to process more than one site in the same crawl. If the seed contains a non-www subdomain ("http://blog.diffbot.com" or "http://support.diffbot.com") Crawlbot will restrict spidering to the specified **subdomain**. If you wish to expand your crawl to multiple domains, enter each one as a separate seed, or consider the `restrictDomain` setting. [Read more.](#)
3. A Diffbot API to be used for processing pages.

Process pages via **ANALYZE API**

Analyze API
http://api.diffbot.com/v3/analyze

Article API
http://api.diffbot.com/v3/article

Image API
http://api.diffbot.com/v3/image

Product API
http://api.diffbot.com/v3/product

Select your Diffbot API from the dropdown menu.

Creating an Automatic Crawl Using the Analyze API

The simplest Crawlbot crawl will apply Diffbot's **Analyze API** to a single site. The Analyze API determines the "page-type" of submitted URLs, and automatically extracts the contents if the page-type is currently supported.

To run your crawl:

1. Enter a new crawl name.
2. Enter a seed URL, like "http://blog.diffbot.com."
3. Select the Analyze API from the "Diffbot API" menu.
4. Click "Start"

In this situation, each page found on the site will be analyzed -- and all supported page-types (article, discussion, image, product, etc.) will be automatically extracted and made available in the resulting collection.

Source: <https://www.diffbot.com/dev/docs/crawl/>

26. In accordance with claim 10 of the '430 patent, Defendant generates instructions for a harvester wherein the harvester provides requests to a server and collects available pages from the server. For example, an analyzer looks for elements on a page and an extractor takes elements out. The harvester collects data from the web pages and put into table based on instructions provided by Diffbot's customers.

Controlling or Limiting Pages to Crawl

Related: What's the Difference Between Crawling and Processing?

By default, Crawlbot will spider and process all links on the domain(s) matching your seed URL(s). You can limit which pages of a site to crawl (spider for links), and which pages of a site to process, using the "Crawling Limits" and "Page Processing Limits" sections.

Crawl Patterns

Crawl patterns will limit crawling to only those URLs containing at least one of the matching strings. You may enter as many crawl patterns as you like, one per line. Any URLs not containing a match will be ignored.

For example, to limit crawling at diffbot.com to only "blog.diffbot.com" pages, you can enter a crawl pattern of `blog.diffbot.com`. If you only wanted to crawl the "Shoes" section of a site, you might enter a crawl pattern of `/category/shoes`.

You can also supply **negative** crawl patterns by prepending your pattern with a "!" (exclamation point). You can supply as many negative patterns as you like. All matching URLs will be skipped.

Crawl Patterns	/category/ !/author/ !/page/
----------------	------------------------------------

This restricts the spider to "category" URLs, and prevents any URLs with "/author/" or "/page/" from being spidered.

You can also use the `^` and `$` characters -- borrowed from regular-expression syntax -- to specify the start and end of a pattern. For instance, `^http://www.diffbot.com` will match URLs starting with "http://www.diffbot.com," and `type=product$` will match all URLs ending in "type=product."

Page Processing Patterns

Page processing patterns are identical to crawling patterns, except they define which pages will be **processed** by the selected Diffbot API.

Note: If you limit site crawling via crawl patterns or regular expressions, but do not enter processing patterns, **only pages matching your crawl patterns/regexes** will be processed. A crawling restriction will serve for both crawling and processing in the absence of a dedicated processing restriction.

For example, to only process product pages for a site you might enter a page processing pattern of

Source: <https://www.diffbot.com/dev/docs/crawl/>

27. The elements described in paragraphs 21-26 are covered by at least claim 10 of the '430 patent.

INFRINGEMENT OF THE '430 PATENT

28. Plaintiff realleges and incorporates by reference the allegations set forth in paragraphs 1 to 27.

29. In violation of 35 U.S.C. § 271, Defendant is now, and has been directly infringing the '430 patent.

30. Defendant has had knowledge of infringement of the '430 patent at least as of the service of the present complaint.

31. Defendant has directly infringed and continues to directly infringe at least claim 10 of the '430 patent by using the Accused Product without authority in the United States, and will continue to do so unless enjoined by this Court. For example, Defendant's website, which is available in Massachusetts, utilizes and implements the Accused Product and performs each and every step recited in claim 10 of the '430 patent.³ As a direct and proximate result of Defendant's direct infringement of the '430 patent, Plaintiff has been and continues to be damaged and harmed.

32. By engaging in the conduct described herein, Defendant has injured SSA and is thus liable for infringement of the '430 patent, pursuant to 35 U.S.C. § 271.

33. Defendant has committed these acts of infringement without license or authorization.

34. As a result of Defendant's infringement of the '430 patent, SSA has suffered monetary damages and is entitled to a monetary judgment in an amount adequate to compensate for Defendant's past infringement, together with interests and costs.

35. SSA will continue to suffer damages in the future unless Defendant's infringing activities are enjoined by this Court. As such, SSA is entitled to compensation for any continuing and/or future infringement up until the date that Defendant is finally and permanently enjoined from further infringement.

³ <https://www.diffbot.com/customers/>.

DEMAND FOR JURY TRIAL

36. SSA demands a trial by jury of any and all causes of action.

PRAYER FOR RELIEF

WHEREFORE, SSA prays for the following relief:

1. That Defendant be adjudged to have infringed the Patent-in-Suit directly, literally and/or under the doctrine of equivalents;

2. That Defendant, its officers, directors, agents, servants, employees, attorneys, affiliates, divisions, branches, parents, and those persons in active concert or participation with any of them, be permanently restrained and enjoined from directly infringing the Patent-in-Suit;

3. An award of damages pursuant to 35 U.S.C. § 284 sufficient to compensate SSA for the Defendant's past infringement and any continuing or future infringement up until the date that Defendant is finally and permanently enjoined from further infringement, including compensatory damages;

4. An assessment of pre-judgment and post-judgment interest and costs against Defendant, together with an award of such interest and costs, in accordance with 35 U.S.C. § 284;

5. That Defendant be directed to pay enhanced damages, including SSA's attorneys' fees incurred in connection with this lawsuit pursuant to 35 U.S.C. § 285; and

6. That SSA have such other and further relief as this Court may deem just and proper.

Dated: July 12, 2017

Respectfully Submitted,

By: /s/ Gustavo A. Chico-Barris

Gustavo A. Chico-Barris

USDC-MA No. 568818

Ferraiuoli LLC

221 Plaza, 5th Floor

221 Ponce de León Avenue

San Juan, PR 00917

Telephone: (787) 766-7000

Facsimile: (787) 766-7001

Email: gchico@ferraiuoli.com

Eugenio J. Torres-Oyola

(Pending pro hac vice admission)

USDC-PR No. 215505

Ferraiuoli LLC

221 Plaza, 5th Floor

221 Ponce de León Avenue

San Juan, PR 00917

Telephone: (787) 766-7000

Facsimile: (787) 766-7001

Email: etorres@ferraiuoli.com

Jean G. Vidal-Font

(Pending pro hac vice admission)

USDC-PR No. 227811

Ferraiuoli LLC

221 Plaza, 5th Floor

221 Ponce de León Avenue

San Juan, PR 00917

Telephone: (787) 766-7000

Facsimile: (787) 766-7001

Email: jvidal@ferraiuoli.com

**ATTORNEYS FOR PLAINTIFF
SSA COMMUNICATIONS LLC**